WHAT IS CLAIMED IS:

1. A method for detecting an information item within an information sequence obtained from a digital medium, said information item comprising any one of a specified set of prestored information items, comprising:

transforming each of said set of prestored information items into a respective representation, in accordance with a predetermined transformation format;

transforming said information sequence obtained from said digital medium, in accordance with said transformation format;

determining the presence of one or more of said prestored information items within said transformed information sequence, utilizing said respective representation.

2. A method according to claim 1, further comprising storing said representations in a database.

3. A method according to claim 1, further comprising sorting said representations into a sorted list.

4. A method according to claim 3, wherein said sorting is in accordance with a tree sorting algorithm.

5. A method according to claim 1, wherein said information item comprises a single word.

6. A method according to claim 1, wherein said information item comprises a sequence of words.

7. A method according to claim 1, wherein said information item comprises a delimited sequence of sub-items.

8. A method according to claim 7, wherein each of said sub-items comprises a sequence of alphanumeric characters.

9. A method according to claim 1, wherein a type of said information item comprises one of a group of types comprising: a word, a phrase, a number, a credit-card number, a social security number, a name, an address, an email address, and an account number.

10. A method according to claim 1, wherein said information sequence is provided over a digital traffic channel.

11. A method according to claim 10, wherein said digital traffic channel comprises one of a group of channels comprising: email, instant messaging, peer-to-peer network, fax, and a local area network.

12. A method according to claim 1, wherein said information sequence comprises the body of an email.

13. A method according to claim 1, wherein said information sequence comprises an email attachment.

14. A method according to claim 1, further comprising retrieving said information sequence from a digital storage medium.

15. A method according to claim 18, wherein said digital storage medium comprises a digital cache memory.

16. A method according to claim 1, wherein said representation depends only on the textual and numeric content of the information item.

17. A method according to claim 1, wherein said transforming comprises Unicode encoding.

18. A method according to claim 1, wherein said transforming comprises converting all characters to upper-case characters or to lower-case characters.

19. A method according to claim 1, wherein said transforming comprises encoding an information item into a numeric representation.

20. A method according to claim 1, further comprising applying a first hashing function to said representations.

21. A method according to claim 1, wherein said information sequence comprises sub-sequences.

22. A method according to claim 21, wherein said sub-sequences are separated by delimiters.

23. A method according to claim 22 wherein said sub-sequences separated by delimiters are any of: words; names, and numbers.

24. A method according to claim 23, further comprising scanning said information sequence to identify said sub-sequences.

25. A method according to claim 24, and said determining is performed by matching said information item to an ordered series of said sub-sequences.

26. A method according to claim 1, further comprising applying a policy upon the detection of said information item in said information sequence.

27. A method according to claim 26, wherein said policy is a security policy, said security policy comprises at least one of the following group of security policies:

blocking said transmission, logging a record of said detection and detection details, and reporting said detection and detection details.

28. A method according to claim 26, wherein said information items are divided into sets, and wherein said security policy depends on the number of detected information items that belong to the same set.

29. A method according to claim 28 wherein each of said sets comprises information items associated with a single individual.

30. A method according to claim 1, wherein said information item comprises a sequence of sub-items.

31. A method according to claim 30, wherein said sub-items are separated by delimiters.

32. A method according to claim 30, wherein a sub-item comprises one of a group comprising: a word, a number, and a character string.

33. A method according to claim 30, wherein said determining comprises using a state machine operable to detect said sequence of delimited sub-items within said information sequence.

34. A method according to claim 30, wherein said transforming comprises:

applying a first hashing function to assign a respective preliminary hash value to each sub-item within said information item; and

applying a second hashing function to assigning a global hash value to said information item based on said preliminary hash values of said sub-items.

35. A method according to claim 34, wherein said information sequence comprises sub-sequences, and wherein said determining comprises:

applying said first hashing function to assign a respective preliminary hash value to each of said sub-sequences;

applying said second hashing function to at least one of said preliminary hash values to assign a global hash value to said at least one of said sub-sequences; and

comparing said global hash value to hash values of said sub-sequences.

36. A method according to claim 35, wherein said sub-sequences comprise one of a group comprising: a word, a number, and a character string

37. A method according to claim 35, wherein said plurality of series comprises a plurality of ordered combinations of sub-sequences within said data sequence.

38. A method according to claim 36, wherein said plurality of series comprises a plurality of combinations of sub-sequences within said data sequence.

39. A method according to claim 38, wherein said second hash function is invariant to reordering of at least two of said sub-sequences.

40. A method according to claim 39 further comprising checking whether said delimited segment was previously stored, and continuing said detection process only if the current delimited segment was previously stored.

41. A method for determining the absence of a specified data item from a list of data items, comprising:

(d) providing an initialized array of indicators;

(e) for each member of said list, performing:

(f) encoding said member with an encoding function to an integer no greater than the size of said array; and

i. setting a corresponding indicator;

ii. encoding said specified data item with said encoding function; and

iii. determining the status of an indicator corresponding to said encoded data item.

42. A method according to claim 41, wherein a size of said array is greater than the number of items in said list.

43. A method according to claim 41 wherein said encoding function comprises a hashing function.

44. A method according to claim 41, wherein a data item comprises a string of alphanumeric characters.

45. A method for determining the absence of a specified data item from a list of data items, comprising:

providing a plurality initialized array of indicators, each of said arrays being associated with a respective encoding function for encoding a data item into an integer no greater than the size of said respective array;

for each of said arrays, performing:

encoding each member of said list with said respective encoding function; and

setting a corresponding indicator for each of said encoded members;

encoding said specified data item with each of said encoding functions; and

for each of said encoded data items, determining the status of the corresponding indicator in said respective array.

46. A method according to claim 45, wherein the size of each of said arrays is greater than the number of items in said list.

47. A method according to claim 45, wherein at least one of said encoding functions comprises a hashing function.

48. A method according to claim 45, wherein a data item comprises a string of alphanumeric characters.

49. An apparatus for detecting an information item within an information sequence, said information item being any one of a specified set of data items, comprising:

a preprocessor, for transforming said information item into a representation, in accordance with a transformation format; and

a scanner, for scanning said information sequence to identify sub-sequences; and

a comparator associated with said preprocessor and said scanner, for comparing said representation to said sub-sequences to determine the presence of said specified information item within said information sequence.

50. An apparatus for detecting a specified information item within an information sequence according to claim 49, further comprising a user interface for inputting said information items.

51. An apparatus for detecting a specified information item within an information sequence according to claim 49, wherein said scanner is further operable to transform said information sequence in accordance with said transformation format.

52. An apparatus for detecting a specified information item within an information sequence according to claim 49, wherein said scanner is further operable to transform said sub-sequences in accordance with said transformation format.

53. An apparatus for detecting a specified information item within an information sequence according to claim 49, further comprising a database for storing a representation of each data item of said set.

54. An apparatus for detecting a specified information item within an information sequence according to claim 49, wherein said information sequence is obtained from a digital medium.

55. An apparatus for detecting a specified information item within an information sequence according to claim 49, further comprising a sorter, for forming a sorted list of the respective representations of set of data items.

56. An apparatus for detecting a specified information item within an information sequence according to claim 49, wherein a type of said information item comprises one of a group of types comprising: a word, a phrase, a number, a credit-card number, a social security number, a name, an address, an email address, and an account number.

57. An apparatus for detecting a specified information item within an information sequence according to claim 49, wherein said information sequence is provided over a digital traffic channel.

58. An apparatus for detecting a specified information item within an information sequence according to claim 49, further comprising retrieving said information sequence from a digital storage medium.

59. An apparatus for detecting a specified information item within an information sequence according to claim 58, wherein said digital storage medium comprises digital storage medium within a proxy server.

60. An apparatus for detecting a specified information item within an information sequence according to claim 49, further comprising a non-existence module comprising:

an encoder, for encoding said sub-sequences and said data item with an encoding function to respective integers, each of said integers being no greater than the size of said array; and

an array setter associated with said encoder, for setting indicators in an array of indicators in accordance with said encoded sub-sequences; and

a status checker associated with said encoder and said array setter, for determining the status of an indicator corresponding to said data item.

61. An apparatus for detecting a specified information item within an information sequence according to claim 49, wherein said encoding function comprises a hashing function.

62. A method according to claim 2, wherein said transforming said representation and storage of said information items comprises:

a) assigning a hash value to each delimited segment within said information item;

b) assigning a hash value for said information item based on said hashes assigned to delimited segments within said information item;

c) storing said hash values evaluated in step a) and step b) above;

and wherein detecting said information items within said digital medium comprises:

d) assigning a hash value to each delimited segment within said digital medium utilizing the same hash function used in step a) above;

e) assigning a hash value for sequences of delimited segments utilizing the same hash function used in step b) above, said sequences being of pluralities of possible numbers of delimited segments within said information items;

f) comparing the hashes values evaluated in step e) above with said hash values stored in step e) above.